

Clinical Research Databases, getting from Current State to Best Practice

Steven E. Waldren, MD, Joseph Quetsch, MD

University of Missouri, Department of Healthcare Management and Informatics,
Columbia, Missouri

Abstract

Clinical research databases reside at the foundation of health sciences, yet experts outside of information science often manage research data. Improper data management presents risks to the researcher, research community, research subjects, and other patients. Data management must aim to preserve or maximize the value of data. We created a checklist tool for researchers and/or data administrators to use when evaluating their database operations. The tool addresses both liabilities and the value of the data.

Background

Databases are commonly utilized in clinical research. Many research databases are maintained outside of a mainstream patient record system¹. This places the management responsibility on the researcher. When researchers are unaware of best practices for handling data, there is risk of unethical conduct of research, violation of privacy ethics, or simply wasted resources or research opportunities. Problems may include: data stores that do not yield their intended information, “bad” data², lost data, data unsuitable for extension or reuse³, data that cannot be validated or audited, data that are unnecessarily difficult to use or permit improper release of private information.

There are many concepts known among information professionals that could be applied by clinical researchers when building, buying or using data stores. These concepts include validating input, controlling terminology, correctly modeling the data, enforced privacy, built-in security, collaboration, and modularization.

Methods

A sample of clinical research databases currently in use at the University of Missouri was analyzed by the authors to qualify and quantify various characteristics believed to be non-trivial for best practice. Items checked included determinants of data integrity, security, privacy, and usefulness. A Sample of questions in the survey tool are in the table.

Results

Three databases were reviewed for initial evaluation of the survey tool. The average time to conduct the survey was 30 minutes. 85% of the 39 questions were able to be answered by direct observation or questioning the database manager. There was a wide range of adoption of the best practice concepts in the survey, no database excelled in all areas. The tool

Data Integrity

Are Standard Code sets used
How is Referential Integrity handled
Where is the data model documented
Can data be audited for Who, When, and What when data is changed
Can primary keys be changed by the user
Is data validated before entry into the database

Security

Levels of access control
How recent is the backup, and has it been tested

Privacy

How is the data de-identified
When is the data de-identified
Is the data store registered with an IRB or Privacy Board

Usefulness

Satisfaction with database
Dates of use & changes in databases or versions

identified and characterized areas for improvement of each data store.

Conclusion

A set of simple questions phrased for any researcher can assess the quality and practices of a research database relative to what professional database administrators or compliance officers consider best practices. The survey tool is suitable for use throughout an organization that is taking an inventory of data stored by researchers. It may be useful for raising expertise in managing clinical research data and assessing educational efforts. The next phase is to refine and validate the survey as a statistical measure of the quality of research data.

Acknowledgements

The National Library of Medicine, grant 2-T15-LM07089-11, supported this work.

References

1. Collen, M.F. (1991). Clinical research databases - A historical review. *Journal of Medical Systems*, 14(6), 323-344.
2. Poloniecki J. Back to databasics. *British J of healthcare Computing & Information Management*. December 1994;11(10):23-4.
3. McDonald CJ, Schadow G, Suico J., Overhage JM. Data standards in health care. *Ann Emerg Med*. September 2001;38:303-11.